REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE
January 31, 2002

3. REPORT TYPE AND DATES COVERED
Final Progress Report 01 Oct 99
~~October 1998- December 2002~~ 30 Sep 02

4. TITLE AND SUBTITLE
Supporting QOS in Mobile Networks

5. FUNDING NUMBERS
DAAD 19-99-1-0287

6. AUTHOR(S)
Professors Andrew T. Campbell and Mischa Schwartz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Columbia University, Office of Projects and Grants
351 Engineering Terrace, New York, NY 10027

8. PERFORMING ORGANIZATION REPORT NUMBER
DAAD 19-99-1-0287

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
39009.1 - CI

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Research includes new "stateful" and "stateless" quality of service (QOS) architectures, services, and algorithms for the delivery of audio, video, and web data over mobile ad hoc networks (MANETs). Major research contributions include the design, development, analysis, and evaluation of three new protocols that address the problem of MANET QOS at different points in the solution space. The *INSIGNIA* protocol represents a new "stateful" QOS architecture that requires per-flow soft-state reservations suitable for multimedia communications are set-up and maintained in the network using in-band signaling techniques. The *SWAN (Stateless Wireless Ad hoc Networks)* protocol represents a departure from per-flow reservations and is based on a stateless approach that cannot offer the same level of QOS performance as INSIGNIA but better scales to support very large numbers of nodes. The *Hotspot Mitigation Protocol (HMP)* interacts with the MANET routing protocols to redirect new "routes" away from hotspot and congestion-prone areas, therefore, avoiding any further build up of traffic intensity in hotspot regions. Each protocol developed by the project is evaluated using a combination of analysis, simulation, and results from experimental wireless testbed deployment. The protocol software from the testbeds and ns simulator extensions is publicly available as open source from the web.

14. SUBJECT TERMS
Quality of Service (QOS), Mobile and Wireless Ad Hoc Networks, Distributed Resource Control

15. NUMBER OF PAGES
30

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

20030306 145

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

1. Table of Contents

2. List of Illustrations and Tables

3. Statement of the Problem Studied

A key challenge to the development of next-generation mobile ad hoc networks (MANETS) is to enable mobile users to access, manipulate and distribute voice, video, data and images with suitable quality of service (QOS). Mobile ad hoc networks are multi-hop in nature, bandwidth limited and are currently designed to support only best effort voice and data communications. Nodes that constitute the wireless network infrastructure are free to move randomly and organize themselves in arbitrary fashions. Thus, the network's wireless topology may be relatively static over periods of time or, more likely, change rapidly in unpredictable ways. Such a network may operate in a standalone fashion, or may be interconnected to intranets or the global Internet via gateway routers. Quality of service support needs to be highly adaptive and responsive to changes in the available resources along the path between two communicating mobile hosts and to network topology changes due to host/router mobility. The objective of this research is to determine the level of quality of service assurances that can be given to mobile hosts/routers in wireless ad hoc networks.

A number of architectural and algorithmic tradeoff exists in designing network support for QOS in MANETs. "Stateful" approaches require that state information is set up and maintained in the network in support of per-flow reservations. However, such approaches may limit scalability and suffer from complexity issues because of the need to introduce and maintain per-flow state information in a mobile ad hoc network. This could become a problem as the number, and the rate of mobility of MANET devices increases. There is a need to study these issues and investigate more scalable and "stateless" approaches to the MANET QOS problem with the goal of seeing how close we can come to achieving the same level of QOS offered by per-flow stateful approaches but without the overhead of maintaining complex state information in the network. The tradeoff between stateful and stateless approaches need to be best understood in terms of pros and cons, with particular emphasis on performance impact.

The project also investigates a number of architectural issues. Can protocol solutions be developed around the idea of a control plane for MANETs where signaling solutions interact with the control protocols (e.g., routing, admission control) in an independent manner, or, is it necessary to embedded control in the routing algorithm itself, as in the case of QOS routing schemes found in wireline research? A benefit of separation would be that any protocols or mechanisms developed by the project would be capable of interworking with a wide range of previously developed MANET routing protocols.

We investigate a set of potential solutions and discuss their tradeoffs. The research methodology used is one of first understanding the problem space, proposing a set of solutions, and then through analysis, simulation and experimentation best understanding the limitations of the solution space. There is an emphasis on experimental systems research where a number of protocols are implemented in experimental MANET testbeds for validation.

4. Summary of the Most Important Results

This project started October 1999 and has produced three new protocols that study the various tradeoffs in the solution space to address the research question discussed in the previous section. Each protocol is fully designed, implemented evaluated. In each case the source code for the protocol is publicly available from the web for experimentation.

We first studied new "stateful" architectures, services and algorithms for the delivery of audio, video and web data over mobile ad hoc networks. Stateful mobile ad hoc networks require that state information is set up and maintained in the network in support of per-flow reservations. In [6] we discussed results from on our work on the INSIGNIA system, which took a stateful approach to building QOS support into mobile ad hoc networks.

Following our work on INSIGNIA we have investigate more scalable and "stateless" approaches to the same problem with the goal of seeing how close we can come to achieving the same level of QOS support without the overhead of maintaining complex state information in the network. We call our new approach Stateless Wireless Ad Hoc Networks (SWAN) [11] and investigate the service differentiation these networks can offer applications.

In the third QOS approach we study new techniques to mitigate "hotspots" using lightweight, local, and scalable algorithms that exploit the routing state maintained at each node in the network. Hotspots represent transient but highly congested regions in wireless ad hoc network that result in increased packet loss, end-to-end delay, and out-of-order packets delivery. In [5] we present a simple, efficient, effective, and scalable *Hotspot Mitigation Protocol (HMP).* Mobile nodes independently monitor their local buffer, contention, and delay conditions, and take local actions in response to emergence of hotspots. HMP balances resource consumption among neighboring nodes, and improves end-to-end throughput, delay, and packet loss.

In what follows, we outline the major research accomplishments for the project. Following this, we provide a brief outline of our research findings for the INSIGNIA, SWAN and HMP protocols.

4.1 Research Accomplishments

New Stateful and Stateless QOS Architectures, Algorithms and Services

We define and evaluate a new IP-based QOS framework for mobile ad hoc networking called INSIGNIA based on a adaptive per-flow soft-state approach. The INSIGNIA framework includes new algorithms for signaling, admission control, packet forwarding, packet scheduling, and medium access. We also define and evaluate new stateless algorithms for mobile ad hoc networking called SWAN. We use rate control for UDP and TCP best-effort traffic, and sender-based admission control for UDP real-time traffic. SWAN uses explicit congestion notification (ECN) to dynamically regulate admitted real-time traffic in the face of network dynamics brought on by mobility or traffic overload conditions. INSIGNIA represents the first in-band signaling approach for pre-flow reservations based on a soft-state approach. SWAN is the first stateless approach for supporting scalable QOS in MANETs.

Hotspot Mitigation Protocol

We develop new techniques to mitigate "hotspots" in MANETs using lightweight, local, and scalable algorithms that exploit the routing state maintained at each node. In [5], we present a simple, efficient, effective, and scalable hotspot mitigation protocol (HMP). Mobile nodes independently monitor their local buffer, contention, and delay conditions, and take local actions in response to emergence of hotspots. HMP balances resource consumption among neighboring nodes, and improves end-to-end throughput, delay, and packet loss. Our results indicate that HMP can improve the network connectivity, preventing premature network partitions.

Results from Analysis and Simulation

We evaluate the performance of INSIGNIA, SWAN and HMP operating with a number of MANET protocols using the *ns* simulator. These routing protocols include Ad hoc On-demand Distance Vector (AODV), Dynamic Source Routing (DSR) and the Temporally Ordered Routing Algorithm (TORA). Extensive evaluation of the protocols indicates the tradeoffs and cost of each protocol for a number of system performance metrics. Results have been widely published [1]-[23] and the ns-2 code made publicly available [21] [22] [23].

Wireless Experimental Testbeds

We built two experimental testbed for INSIGNIA and SWAN based on IEEE 802.11. The INSIGNA testbed [21] implements the protocol with the DSR routing protocol. The testbed was successful publicly demonstrated (www.argreenhouse.com/ mobicom2000/Demos.htm) at the ACM Sixth Annual International Conference on Mobile Computing and Networking *(ACM MobiCom 2000)*, Boston, August 6-11, 2000. The SWAN testbed [22] implements the protocol with the AODV routing protocol.

IETF Contributions and Impact

We presented the INSIGNIA signaling system within the IETF MANET Working Group. The INSIGNIA signaling specification [17] is published as an official working group Internet Draft (www.ietf.org/html.charters/manet-charter.html). We submitted an Internet Draft on SWAN [18] to the IETF Working Group on MANETs. We are in the process of preparing a new Internet Draft on the Hotspot Mitigation Protocol for submission to the MANET Working Group at the next meeting (56th IETF) in San Francisco, CA, USA (March 16-21, 2003). While the issue of QOS is still premature for the working group's agenda we have been major contributors to new ideas, techniques, and protocols that could be used in MANETs.

Open Source Code Release and Impact

We have the led the way in releasing open source releases for the protocols developed in the ARO project [21] [22] [23]. We first released the INSIGNIA signaling system as extensions to the *ns* simulator in August 2000. The code interoperates with the main routing protocols discussed in the MANET Working Group (i.e., AODV, DSR and TORA) and can be downloaded from the INSIGNIA Project home page [21]. In June 2001 we released the testbed open source code for INSIGNIA. The testbed software [21] has seen over 700 downloads since that date with a number of industrial and academic laboratories building enhancements to the testbed code (including Ericsson Research, Nortel Networks, Hughes). Recently, we released the SWAN software [22] as extensions to the *ns* simulator and as a full testbed source code release in October 2002. The SWAN software [22] has seen over 130 downloads since its release.. The ns-2 SWAN code interoperates with the main routing protocols discussed in the MANET Working Group (i.e., AODV, DSR and TORA) and the testbed software is released with the AODV stack. We are currently working on the software release [23] for HMP for summer 2003.

Publications

There has been a significant number of publications [1]-[23] from the project: *i)* six journal papers [6]-[11] have been published in the best IEEE and ACM journals. For example, *IEEE Transactions of Mobile Computing* published our SWAN paper [11] as one of the top 6 wireless papers to be published in IEEE INFOCOM 2002. Over 900 papers were submitted to INFOCOM 2002; *ii)* five conference papers [12]-[16] were published; and *iii)* two Internet Drafts for INSIGNIA [17] and SWAN [18]; and *iv)* two open source code releases for the INSIGNIA [21] and SWAN [22] testbeds. All papers and source code is online [21] [22] [23].

4.2 Research Findings

In [4] we specify a new IP-based QOS framework that supports differentiated and adaptive services in mobile ad hoc networks. Architecturally the INSIGNIA QOS framework is designed to support fast reservation, restoration and end-to-end QOS adaptation based on the inherent flexibility, robustness and scalability found in IP networks. We evaluate the framework paying particular attention to the performance of the in-band signaling system, which helps counter time-varying network dynamics in support of the delivery of adaptive services. Our results show the benefit of our framework to deliver lower end-to-end delay and enhanced throughput under diverse mobility, traffic and channel conditions.

In [1] we propose SWAN, a stateless network model which uses distributed control algorithms [15] [10] to deliver service differentiation in mobile wireless ad hoc networks in a simple, scalable and robust manner. We use rate control for UDP and TCP best-effort traffic, and sender-based admission control for UDP real-time traffic. SWAN uses explicit congestion notification (ECN) to dynamically regulate admitted real-time traffic in the face of network dynamics brought on by mobility or traffic overload conditions. We use the term "soft" real-time services to indicate that real-time sessions could be regulated or dropped due to mobility or excessive traffic overloading at mobile wireless routers. SWAN is designed to limit such conditions, however. A novel aspect of SWAN is that it does not require the support of a QOS-capable MAC. Rather, soft real-time services are built using existing best effort wireless MAC technology. Simulation, analysis, and results from an experimental wireless testbed show that real-time applications experience low and stable delays under various multi-hop, traffic and mobility conditions.

The simple goal of HMP is to redirect new "routes" (i.e., the establishment of new routes and therefore the forwarding of data packets along those routes) away from hotspots. HMP disperses new flows away from being routed through hotspots and congestion-prone areas, avoiding the further build up of traffic load in hotspot regions. HMP effectively mitigates hotspot conditions and reduces congestion-related problems. Mitigating hotspot in this manner also fosters balanced resource consumptions among neighboring nodes, and can extend lifetimes of certain overtaxed nodes. Although, the idea behind HMP is rather simple, implementing HMP requires the development of fairly sophisticated mechanisms. First, HMP needs to accurately identify hotspots through local and distributed information. HMP utilizes MAC-delay measurements, buffer occupancy information, neighbor status information and other resource monitoring mechanisms (i.e., buffer, power) to detect hotspots.

In what follows, we present the main research findings from each protocol.

4.2.1 INSIGNIA: A Stateful QOS Architecture

In [6] we present an evaluation of the INSIGNIA QOS framework through simulations with emphasis on the performance of the signaling system. Figure 1 shows that INSIGNIA supports relatively constant QOS under slow and moderate mobility conditions between 3.6-18 km/hr. The optimal performance is observed when the average network mobility is approximately 11 km/hr. This results in the delivery of 86% of reserved packets (i.e., packets with reservations). The in-band nature of INSIGNIA allows the system to cope with fast network dynamics in a responsive manner. In an ideal case, INSIGNIA only requires a single packet reception to restore reservations for re-routed flows. INSIGNIA supports the delivery of 66% of reserved packets even when mobile hosts are moving at 72 km/hr, as shown in Figure 1. This is a very encouraging result.

INSIGNIA represents a general purpose solution to service differentiation in mobile ad hoc networks; that is, INSIGNIA support "operational transparency" among multiple routing protocols through the separation of routing, signaling and packet forwarding. This is in contrast with other approaches found in the literature that call for more integration of resource management and routing to deliver end-to-end QOS. These approaches, however, limit operational transparency through the integration of QOS and routing. For a detailed description of the ISIGNIA model see [] and for a detailed discussion of the results of the protocol see [8] [6].

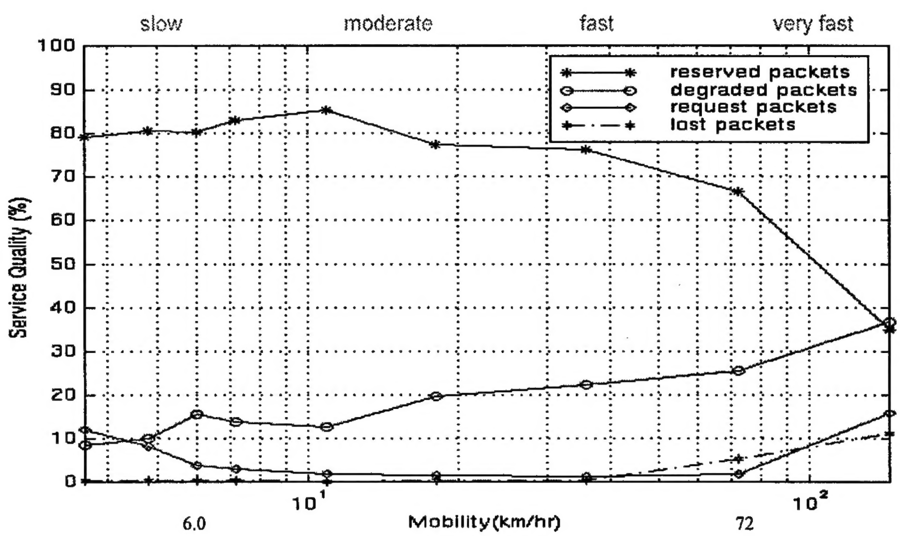
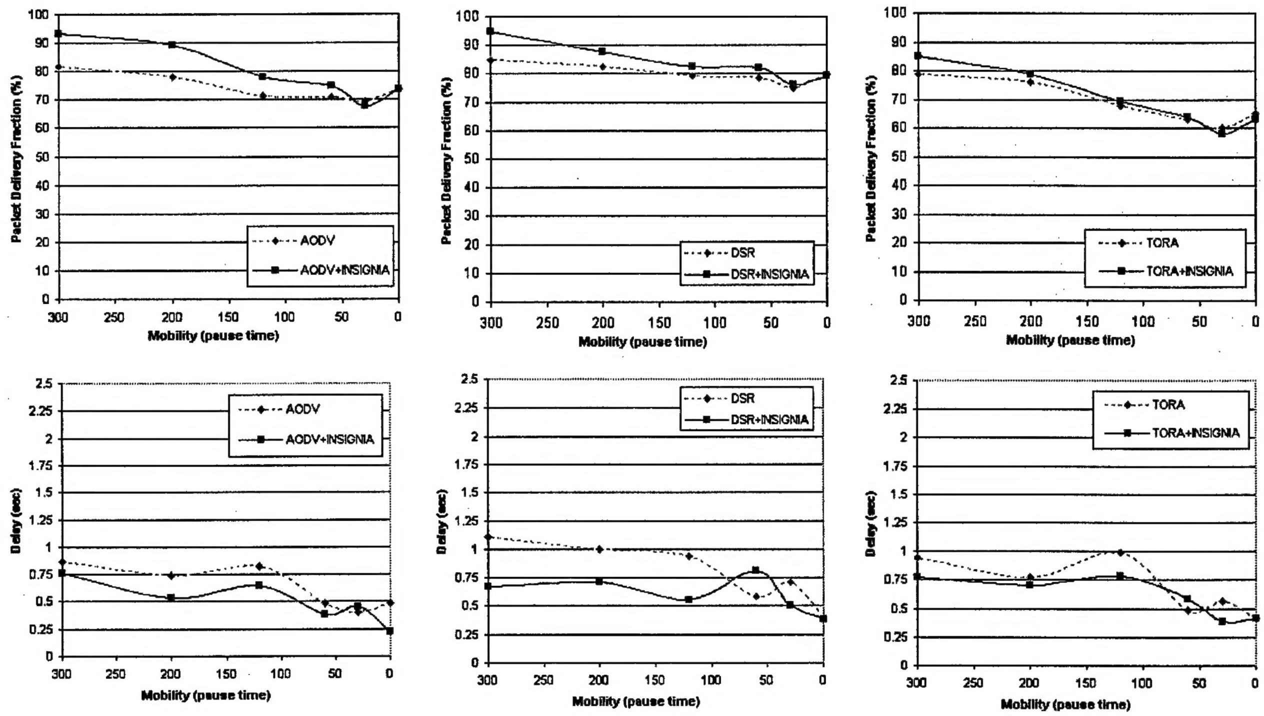


Figure 1. Mobility and Network Performance

4.2.2 TCP and UDP Performance

In [8] we evaluate the DSR, AODV and TORA routing protocols with and without the INSIGNIA signaling system in support of UDP and TCP traffic. Results indicate that INSIGNIA has good operational transparency features where the system provides enhanced end-to-end throughput and lower measured end-to-end delays than the best effort system. The impact of mobility on the packet delivery fraction and measured delay in support of UDP performance is shown in Figure 2. INSIGNIA outperforms the baseline best effort system in terms of the packet delivery fraction

and average end-to-end delay. The INSIGNIA system experiences a 20% increase in delivered packets over the baseline system for moderate to low mobility conditions. End-to-end delay shows the same trend. Mobility has little impact, however, on the average end-to-end delay for the INSIGNIA system. This is not the case for the best effort system where mobility has a greater effect on the measured performance.



(a) AODV (b) DSR (c) TORA

Figure 2. Impact of Mobility with and without INSIGNIA on UDP Performance

The impact of increasing host mobility and traffic load on TCP-Reno, TCP-SACK, TCP-Vegas and TCP-ELFN is illustrated in Figures 3 and 4, respectively. Figure 3 illustrates the impact of mobility on TCP flows in terms of "goodput". Substantial improvement in goodput for the INSIGNIA system is observed at lower mobility conditions where routes are more stable and end-to-end reservation remains stable for longer periods of time. As mobility increases, the improvement of the INSIGNIA system over the best effort system narrows, as shown in Figure 3. The INSIGNIA system not only improves goodput but also provides better service quality under all mobility conditions. At high mobility, TCP flows often decrease their window segment size to the minimum due to packet losses resulting from loss of connectivity or congestion.

We observe that, as the traffic load increases, the measured goodput decreases substantially in all cases. For all TCPs in the best effort system goodput decreases by approximately 60 % when the network load increases. However, the impact of the traffic load on goodput measurements in the INSIGNIA system is much less than that of the best effort system. As observed in Figure 4, the performance differences between the best effort and INSIGNIA systems widen as the network load increases. The QOS improvements become more apparent under heavily loaded network conditions where goodput improvement is more than 90% for all TCPs. For full details of the results of this work see [8].

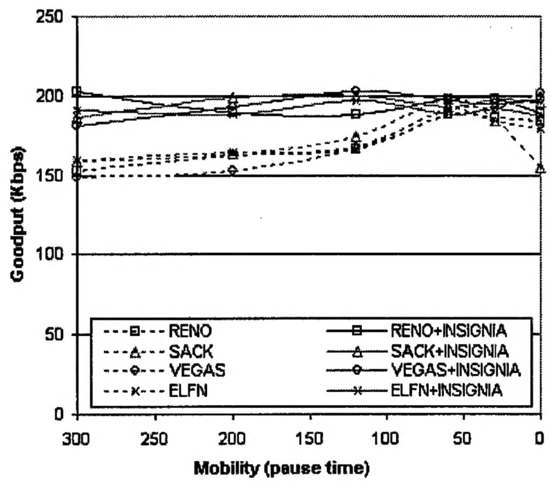


Figure 3. Impact of Mobility on TCP Goodput

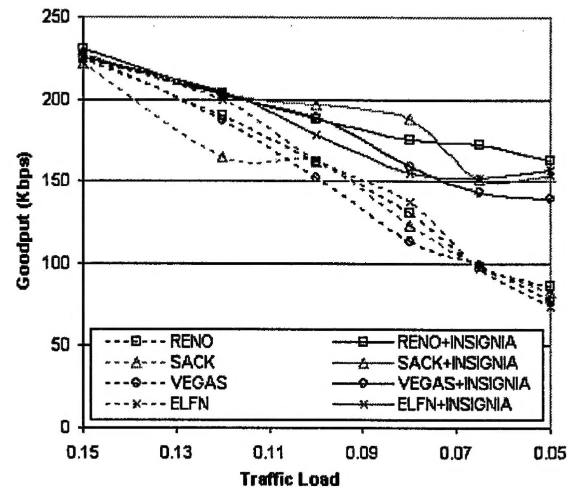


Figure 4. Impact of Load on TCP Goodput

4.2.3 Utility-Fair Adaptive Services and Algorithms

In [9] we propose a number of enhancements to the data link layer architecture to support QOS adaptation for wireless services classes using a utility-fair resource allocation algorithm. A centralized adaptation controller and distributed adaptation handlers form a combined centralized/distributed architecture. The adaptation controller interacts with a bandwidth allocation algorithm for the distribution of bandwidth among competing flows. The bandwidth allocation algorithm is based on the utility curves supplied during flow setup and takes into account wireless service classes selected by the application. The adaptive controller is also responsible for controlling the minimum flow specific adaptation time-scale for active adaptation services. Slower adaptation time-scales can be built on this but are implemented by the adaptation handlers at the mobile device.

A signaling interface (S-MAC) provides an in-band channel between mobile devices supporting various signaling features, e.g., flow setup, adaptation control, and scheduler polling. The S-MAC interface provides a fast signaling channel, which allows response times in the order of a few milliseconds. Specifically, the S-MAC interface supports signaling between the adaptation controller, which determines flow-specific bandwidth allocations, and the distributed adaptation handlers, which are periodically informed of the advertised bandwidth allocation.

In contrast to the centralized nature of bandwidth allocation the adaptation handlers are responsible for enforcing application-specific adaptation behavior for active and passive adaptation services. In the case of active adaptation services, adaptation handlers determine whether or not the application will adapt to any portion of the advertised available bandwidth. This is based on an adaptation policy programmed into the handlers by the application. In this sense the adaptation handler acts as a proxy on behalf of the application to implement policy. This includes the adaptation time-scale over which an application can adapt and the amount of additional bandwidth the application requires before it will adapt. Typically the adaptation controller allocates bandwidth to a flow based on its utility curve and the adaptation handler decides whether to accept the advertised rate or not based on the programmed policy. This interaction between the controller and handlers is only for active adaptation services. Adaptation

handlers are flexible programmable objects that can implement sophisticated adaptation policies (e.g., local buffer management schemes, selective packet discarding techniques, adaptation time-scales, etc.) in addition to bandwidth allocation dynamics.

Bandwidth allocated to adaptation classes is based on a utility-fair share of best effort wireless resources, rather than on explicit application level reservation. When allocating bandwidth preference is given to requests for the active adaptation wireless class over the passive class.

Wireless classes are based on the best effort paradigm while the sustained rate class uses explicit reservation and admission control. In [9] we define bandwidth-based utility curves, and present our bandwidth allocation algorithm and details on the implementation of the proposed scheme. We propose the use of flow-specific bandwidth utility curves that quantitatively model an application's subjective quality in relation to time-varying bandwidth conditions that a mobile device may experience over the air-interface.

In [9] we define a bandwidth utility curve as a mapping of the available wireless network bandwidth into a utility at any time. Typically, an increase in bandwidth does not decrease the application's quality making the utility curve a non-decreasing function of bandwidth. The satisfaction (i.e., "quality") value that represents the level of satisfaction perceived by an application measure at the application layer is usually coarse and obtained via subjective testing, such as the 5-level mean-opinion-score (MOS) measure for video quality. We characterize utility curves using a limited set of parameters. The utility curve is used for admission control and is signalled across the S-MAC channel at flow setup and during application level renegotiation. Our implementation [9] defines a bandwidth utility curve as a piecewise linear function using a quantized set of utility levels.

We are currently investigating admission control strategies using the techniques discussed in [9] and their application to decentralized control. For full details of the results of this work see [9].

4.2.4 SWAN: A New Stateless Approach to QOS

The SWAN model [16] [11] [20] [21] includes a number of mechanisms used to support rate regulation of best effort traffic. A classifier and a shaper operate between the IP and MAC layers. The classifier is capable of differentiating real-time and best effort packets, forcing the shaper to process best effort packets but not real-time packets. The shaper represents a simple leaky bucket traffic shaper. The goal of the shaper is to delay best effort packets in conformance with the rate calculated by the rate controller.

There is no flow or session state information maintained at intermediate nodes in support of end-to-end communications between source-destination pairs – hence the term "stateless". Furthermore, when a session is admitted there is no admission control decision taken at intermediate nodes. Rather, the admission control test to determine if a new session should be admitted or not is conducted solely at the source node. A key operation of the admission controller, which is based at every mobile device, is to efficiently estimate local bandwidth availability. The admission controller located at the source node probes the network between the source and destination to determine the instantaneous end-to-end bandwidth availability. Based on the results of a *request/response* probe the session admission controller located at source node makes a decision to admit a new real-time flow or not. Once a session is admitted as a real-time session its packets are marked as RT (for real time service) otherwise they are considered as best effort packets. We use the DS (DiffServ) code word to maintain this packet state information in our SWAN wireless testbed and ns-2 simulation environment. Typically, a bandwidth probe is

sent at the beginning of a session or, as discussed later, when mobility or channel load conditions force a real-time session to re-establish its end-to-end service quality.

Once a session is admitted it is desirable to maintain service quality for the lifetime of the session. Because our wireless network model takes a conservative approach when allocating bandwidth to real-time traffic, small scale violations of service quality can be tolerated without impacting application level QOS. These small-scale violations may occur because of bursty real-time sources or unpredictable traffic patterns. In a static wireless ad hoc network there is little need for further control algorithms above and beyond the rate control of best effort traffic and admission control of real-time traffic. One could even argue that under low mobility conditions this approach is sufficient for the delivery of real-time performance. However, the bandwidth availability and dynamics of a wireless channel may change rapidly in the case of moderate to higher levels of mobility. Larger-scale violations may occur when real-time flows are admitted or dynamically re-routed. In the former case, multiple source nodes could simultaneously send new session probes that may traverse common intermediate nodes facilitating the admission of new sessions. This in turn could overload these common intermediate nodes. There is a need for additional SWAN mechanisms that can help resolve these issues.

We regulate real-time sessions when a mobile node observes violation of real-time sessions; for example, due to mobility or source-based admission control. We adopt a regulation mechanism based on ECN, which was originally proposed for controlling and improving TCP traffic performance in IP networks. To allow for experimentation with IPv4, two bits (i.e., ECN-Capable Transport and Congestion Experienced bits) have been set-aside in the IP header for ECN. We use ECN to control and regulate UDP real-time traffic in the case of traffic violations most likely brought on by the re-routing of real-time sessions. By regulation, we mean that the ECN mechanism forces real-time flows to re-establish their real-time service. Under such conditions an existing flow would either be able to re-establish its original service quality or be dropped. We do not consider bandwidth adaptation of real-time sessions as in the case of INSIGNIA. Rather, we assume that real-time flows attempt to re-establish service at their original bandwidth levels. For full details on the results of this work see [6] [8] [9] [12] [13] [17] [19] [21].

4.2.4.1 SWAN Performance

In [16] [11] we analyze the MAC delay and the probability that mobile devices find themselves in a backlogged state in IEEE 802.11 wireless networks. (We use the terms "original system" and "proposed system" to refer to IEEE 802.11 wireless networks with and without rate control, respectively). We show through analysis that the proposed system performs better than the original system in terms of MAC delay. In [11] we explains why this is the case. We show that by controlling the probability of mobile nodes being in a backlogged state, the target MAC delay of the real-time traffic can be maintained. This result confirms that the SWAN approach is feasible and effective.

We implemented SWAN [22] using the ns-2 simulator and its wireless extensions developed at CMU. The SWAN ns-2 extensions include the AIMD rate controller, admission controller, packet delay measurement mechanism, local utilization monitoring, probe protocol for bandwidth availability estimation, and explicit congestion notification.

AIMD Parameter (c, r) Analysis

To better understand the properties of the SWAN AIMD rate control parameters [11] c and r, we consider two scenarios for background TCP best-effort traffic. The first scenario has 8 TCP flows and the second has 32 TCP flows. In both scenarios, all TCP flows are greedy FTP type of traffic with packet size of 512 bytes. TCP flows are rate controlled with parameter c and parameter r, while voice and video flows are not rate controlled once admitted through the source-based admission control process. During the simulation, 4 voice flows and 4 video flows are active and monitored for the duration of 200 seconds representing real-time traffic. Voice traffic is modeled as 32 Kbps constant rate traffic with a packet size of 80 bytes. Video traffic is modeled as 200 Kbps constant rate traffic with a packet size of 512 bytes.

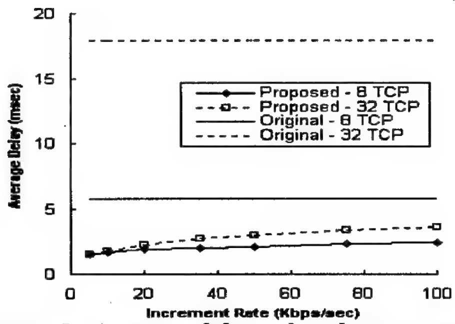


Fig. 5. Average delay of real-time traffic vs. increment rate.

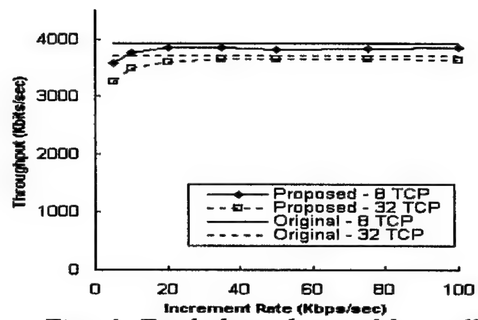


Fig. 6. Total throughput of best-effort TCP traffic vs. increment rate.

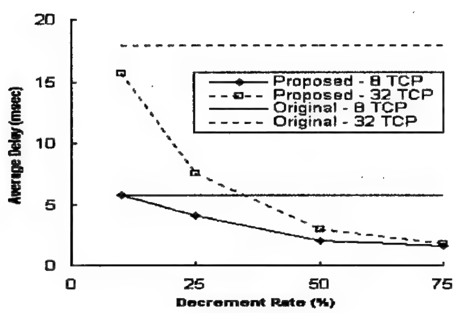


Fig. 7. Average delay of real-time traffic vs. decrement rate.

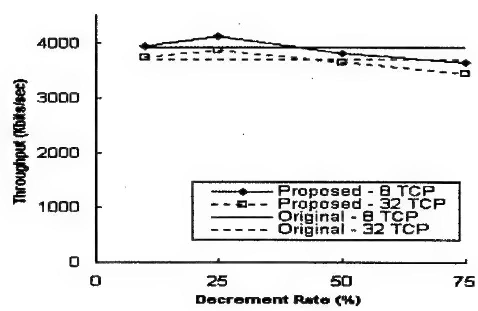


Fig. 8. Total throughput of best-effort TCP traffic vs. decrement rate.

We measured the average MAC delay of real-time traffic (see Figures 5 and 7) and the total throughput of best-effort traffic (see Figures 6 and 8). The x-axis of Figures 5 and 6 represent the value for parameter c (increment rate, Kbps/sec). The x-axis in Figures 7 and 8 represents the value for parameter r (decrement rate, %). It is shown in Figure 5 that the value of parameter c does not have much impact on the average delay of real-time traffic. The average delay grows very slowly with the increasing value of parameter c. In contrast, the total throughput of best-effort TCP traffic is noticeably decreased when a small value of parameter c is chosen, as shown in Figure 6. When the increment rate is 5 Kbps/sec, throughput is reduced by about 10% for the 8 TCP flow scenario and by 13% for the 32 TCP flow scenario in comparison to the original system. For an increment rate of 20 Kbps/sec or larger, the TCP throughput becomes almost constant with less than 3% reduction in throughput.

Performance of Multi-hop Scenarios with Mobility

Figures 5 and 6 show the average end-to-end delay for real-time traffic and TCP best effort traffic for an increasing number of background TCP traffic, respectively. We observe that the packet loss of the real-time traffic is less than 1% in both the original and proposed systems. However, the average delay of the real-time traffic shows a significant difference between the original and proposed systems. The average end-to-end delay of the real-time traffic in the original system grows linearly from 8 to 30 msec, as the number of TCP flows increase from 2 to 12 flows, respectively. In contrast, the average delay of real-time traffic in the proposed system remains around 5 to 7 msec. The average "goodput" of TCP traffic in the proposed system is about 15-20% less than the original system. By adopting the proposed control mechanisms, we observe a 38-77% reduction in the average delay of the real-time traffic at a cost of 15-20% loss of TCP goodput. In addition, the average delay of the real-time traffic remains consistently below 8 msec in the proposed system while the average delay in the original system grows above 30 msec.

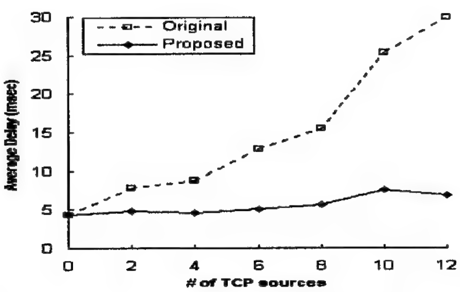


Fig. 9. Average delay of real-time traffic vs. number of TCP flows.

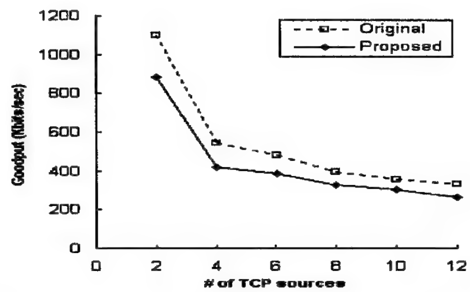


Fig. 10. Average "goodput" of TCP best-effort traffic vs. number of TCP flows.

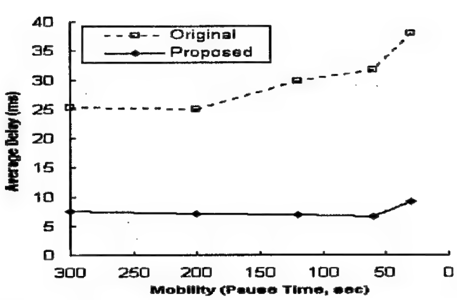


Fig. 11. Average delay of the real-time traffic vs. mobility.

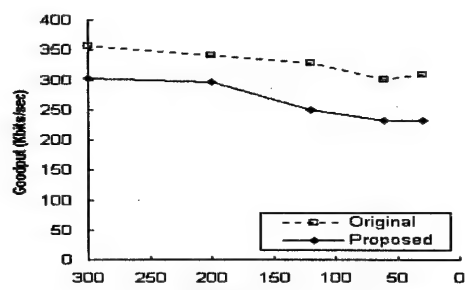


Fig. 12. Average goodput of the best–effort TCP traffic vs. mobility.

The impact of mobility is illustrated in Figures 11 and 12. The simulated network is the same as the previous multi-hop scenarios with the addition of the introduction of mobility. We use a random waypoint mobility model. Each mobile node selects a random destination and moves with a random speed up to a maximum speed of 72 km/hr and pauses for a given "pause time" when the destination is reached. When the pause timer expires, the mobile node picks another random

destination and moves at another random speed. The real-time traffic is modeled in the same manner as discussed previously. The number of best-effort TCP flows comprises 5 FTPs and 5 web micro-flows.

4.2.4.2 Experimental Wireless Testbed

We have implemented SWAN in an experimental wireless testbed based on Linux notebooks using Aironet IEEE 802.11b wireless interfaces. The rate controller is implemented by modifying the Aironet device driver. We also modified the device driver to measure packet delay. The packet delay is measured by calculating the difference between the time the device driver feeds a new packet into an Aironet card and the time the Aironet card acknowledges back to the device driver that the transmission of the packet was successful. We implemented a traffic shaper driver between the kernel and the Aironet card device driver to control the rate of TCP traffic.

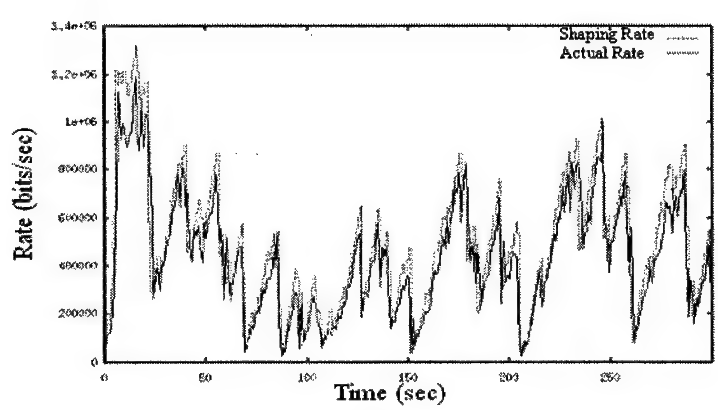


Fig 13. Trace of the shaping rate and the actual TCP transmission rate.

The utilization monitor and probe protocol are implemented using the Berkeley Packet Filter's Packet Capture library (PCAP). PCAP is designed to capture packets for statistical purposes but it can also be used to forward packets to the network interface. PCAP is used to capture every UDP packet transmitted within the radio contact range of a wireless mobile host. The admission controller reads the IP header of captured UDP packets and estimates the local bandwidth availability. We also use PCAP to capture and forward probe signals. The admission controller estimates the end-to-end bandwidth availability when a source node probes the network path, as previously discussed.

The results presented in this section were obtained from a SWAN wireless ad hoc testbed, which consists of five mobile hosts using Aironet 11 Mbps IEEE 802.11b PCMCIA cards. The configuration of the testbed is as follows. Four mobile hosts generated TCP traffic and one mobile host generated UDP traffic. The source and the destination nodes associated with each flow were distributed among the mobile hosts. The UDP host generated packets every 20 msec at 32 Kbps with the rate of the TCP flows being controlled by the rate controller.

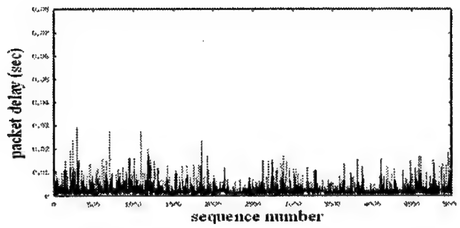


Fig. 14. Delay of each packet in a UDP real-time flow from the SWAN wireless testbed with rate control.

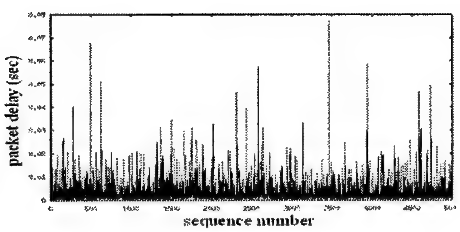


Fig. 15. Delay of each packet in a UDP real-time flow from the SWAN wireless testbed without rate control.

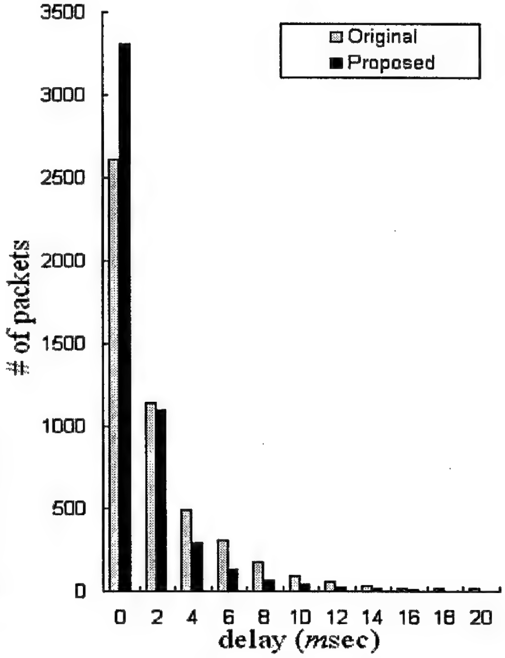


Fig. 16. The distribution of packet delay.

Figure 13 shows a trace of the shaping rate controlled by the rate controller and the actual TCP transmission rate. The actual TCP rate is well controlled by the shaper, as shown in Figure 13. When all four TCP flows were rate controlled, we measured the delay of each packet in a UDP real-time flow. Figures 14 and 15 show the delay of each packet when the TCP flows are regulated and unregulated, respectively. By comparing Figures 14 and 15, we can observe that the measured delay is improved when TCP flows are rate controlled. The average measured delay is 2.3 msec and 3.3 msec in Figure 14 and 15, respectively. The measured delay in Figure 14 remains below a certain boundary most of the time, while the delay in Figure 15 reaches significantly higher values. Figure 16 shows a simple distribution of the measured UDP real-time packet delay for the original (i.e., the wireless testbed without rate control) and proposed systems (i.e., the wireless testbed with rate control). We can observe that proposed system has more packets with packet delays smaller than 2 msec, and fewer packets with measured delays exceeding 4 msec. 66% of the packets have delays of less than 2 msec in proposed system, in comparison to 52% for the original system. In the proposed system only 11% of the packets have delays greater than 4 msec in comparison to 24% for the original system. For full details on the results of this work see [11] [10] [16].

4.2.5 Hotspot Mitigation Protocol

In the third approach to QOS in MANETs we focus on new techniques to mitigate "hotspots" using lightweight, local, and scalable algorithms that exploit the routing state maintained at each node. Hotspots represent transient but highly congested regions in wireless ad hoc network that result in increased packet loss, end-to-end delay, and out-of-order packets delivery. In [5], we present a simple, efficient, effective, and scalable hotspot mitigation protocol (HMP). Mobile nodes independently monitor their local buffer, contention, and delay conditions, and take local actions in response to emergence of hotspots. HMP balances resource consumption among neighboring nodes, and improves end-to-end throughput, delay, and packet loss. Our initial results [5] indicate that HMP can improve the network connectivity, preventing premature network partitions. In what follows, we present a summary of the analysis of hotspots, and the design and evaluation of HMP. We evaluate the protocol's ability to effectively mitigate hotspots in mobile ad hoc networks that are based on on-demand routing protocols, such as, AODV and DSR. For full details on the results of this work see [5].

4.2.5.1 Existence of Hotspots

Figure 17 illustrates typical hotspot conditions found in mobile ad hoc networks. Hotspots are generally created where traffic loads converge to a node or small cluster of nodes. Flows traversing multiple wireless hops from various locations intersect each other and create transient hotspot conditions. We observe that hotspot nodes and nodes in the vicinity of the hotspots are more prone to consume more resources than others. Left unchecked such unbalances resource consumption is detrimental to mobile ad hoc networks because overtaxed nodes would prematurely exhaust their power reserves before other nodes. As a consequence the network connectivity is threatened. In addition, we observe that hotspot nodes are often responsible for generating a large amount of routing overhead. In general, as traffic load increases more hotspots appear and conditions in hotspot prone regions become aggravated.

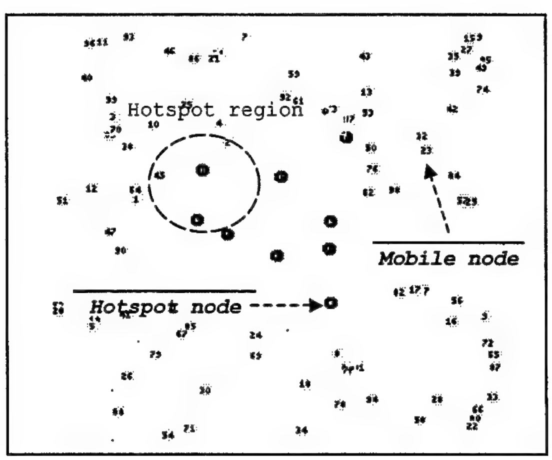


Figure 17. Hotspots Snapshot for AODV-based MANET

Figure 17 is a snapshot of a typical mobile ad hoc network captured during a simulation run using ns-2. The snapshot illustrates conditions in a mobile ad hoc network at its individual nodes at t = 140 seconds. Our simulation consists of 100 mobile nodes in 1200 meter by 1200 meter sized network with moderate mobility condition (i.e., pause time of 80 seconds in random waypoint mobility model). Thirty CBR/UDP and 10 TCP flows are used to produce an offered load of approximately 480 Kbps. Nine hotspots are identified in the snapshot shown in Figure 17. Note that hotspots are often momentary because of the mobility of nodes changes the topology and continuously varies the traffic loads in the network causing hotspots to migrate. We observed that nodes in our simulations are rarely in a permanent hotspot state. Rather, a particular node may come in and out of hotspot conditions. As a rule of thumb we defined a hotspot as a node that remained in a congested low performance mode for a minimum of 5 seconds. Thus under simulation nodes could be declared hotspot a number of times (e.g., 60 times in a particular simulation run). Using this time-scale, we observed 816 congestion hotspot incidents during the simulation run described above where the offered load is only 480 Kbps. Note, that 816 hotspots instances corresponds to a minimum of 4080 seconds of congested conditions, or, an average of 40.8 seconds of congestions per node.

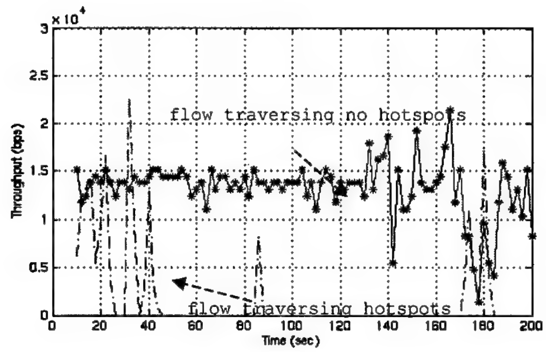


Figure 18 Throughput Traces of Two Monitored Flows

Figure 18 simply illustrates the throughput traces of two similar flows under the simulation configuration discussed above. We selected a flow traversing multiple hotspots, and a flow encountering a single hotspot (from our simulation results) and compare their throughput performance. The trace demonstrates how hotspots impact flow performance. Under the same condition, we also observe the packet loss of nodes in hotspot regions (i.e., hotspot nodes and their single hop immediate neighbors) and compare it to nodes that do not experience hotspots. For full details see [5].

Based on these observations [5] we argue that there is a need to study, design, and evaluate mechanisms that can seamlessly interwork with existing MANET routing protocols to mitigate the impact of hotspots. This approach requires that no QOS state information be maintained in the network, as is the case with INSIGNIA, nor does it require any probing and ECN mechanisms as the case with the SWAN stateless approach. The solution in this case is based on local hotspot mitigation techniques that interact with the distributed routing state. MHP represents a different technique to the stateful and stateless approaches discussed earlier in this report.

4.2.5.2 Hotspot Mitigation Protocol (HMP)

The simple goal of HMP is to redirect new "routes" (i.e., the establishment of new routes and therefore the forwarding of data packets along those routes) away from hotspots. HMP disperses new flows away from being routed through hotspots and congestion-prone areas, avoiding the further build up of traffic load in hotspot regions. HMP effectively mitigates hotspot conditions and reduces congestion-related problems. Mitigating hotspot in this manner also fosters balanced resource consumptions among neighboring nodes, and can extend lifetimes of certain overtaxed nodes.

Although, the idea behind HMP is rather simple, implementing HMP requires the development of fairly sophisticated mechanisms. First, HMP needs to accurately identify hotspots through local and distributed information. As mentioned previously, HMP utilizes MAC-delay measurements, buffer occupancy information, neighbor status information and other resource monitoring mechanisms (i.e., buffer, power) to detect hotspots. HMP does not limit the scope of monitoring or detection mechanisms, however. Operators are free to introduce additional mechanisms and algorithms according to their needs.

HMP utilizes monitored and measured information to respond to conditions by executing the most appropriate algorithms to alleviate the condition at hand. The monitored and measured condition is explicitly expressed by a multimetric parameter called `STATUS`. Status consists of two components: *symptom* and *severity*. Symptom describes the dominant condition a node is experiencing while severity expresses the degree of the symptom. For example, a node may declare its status as $Y_{CONGESTION}$ while another node may declare its status as R_{POWER}. This status is analogous to traffic lights, where green (i.e., denoted by G) indicates a good condition, yellow (Y) represents a marginal condition, and red (R) represents a critical condition. Therefore, $Y_{CONGETION}$ indicates marginal congestion and R_{POWER} indicates critically low power. Users/operators are free to introduce more granularity if needed. HMP piggybacks this status information in the IP option field and neighboring nodes operating in promiscuous mode learn the status of transmitters by eavesdropping their packets. The eavesdropped information is used to create and update a *neighborhood status table (NST).* The cached information is locally maintained and updated at each node.

An NST caches a list of immediate neighbors and their status. It is primarily utilized to manipulate new-route-creation decisions at nodes. In other words, a node refers to its NST to ensure that it is not aggravating the conditions of neighboring nodes by creating additional routes through them. We assume that there is finite number of neighboring nodes surrounding a node and the number of neighboring nodes defines the size of the NST of a node. A node maintains and updates its NST every time it hears a packet. Another role of an NST lies in its neighborhood selection algorithm. A node continuously monitors its NST and identifies bad neighbors. A node is considered to be a 'bad' neighbor when packets are continuously retransmitted in the absence of congestion. If such event occurs, HMP temporarily marks this node as a bad neighbor for $T_{BAD\text{-}NEIGHBOR}$ period and avoids creating new routes through the node. This is achieved by discarding route request packets from a bad neighbor.

However, a naïve suppression of new route creation may prevent the use of the only possible path between two hosts and may yield poor connectivity, or even cause network partitions. To avoid this, a new-route-suppression mechanisms is used, if, and only if, there exists a sufficient number of non-hotspot neighbors within its transmission range. HMP also makes sure that preceding nodes en-route also have enough non-hotspot neighbors. The notion of 'enough neighbors' is defined by `enough_nh_neighbor` parameter (i.e., currently set at 6 in our implementation) where choice of this parameter has an impact on the network connectivity. If `enough_nh_neighbor` is too small say 2 then HMP manifests low connectivity among

mobile nodes and often fails to provide useful routes. HMP also ensures that it is not inadvertently denying the only possible path between two end hosts by utilizing an indicator called `path_indicator`. A node that has only a few neighboring nodes sets this indicator and subsequent nodes en-route acknowledging the indicator avoid suppressing new route creation. This is illustrated in Figure 19 where hotspot M_4 forwards RREQ toward M_5 because source node M_3 has set its path indicator whereas hotspot M_2 suppresses RREQ from M_1 because its `path_indicator` is not set.

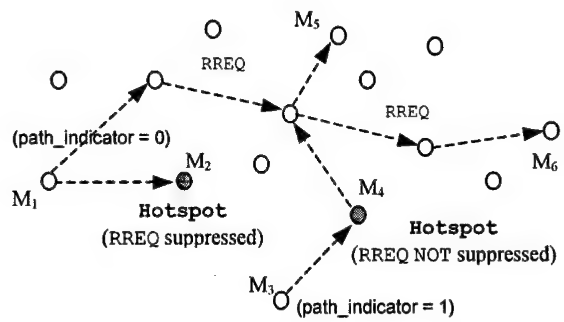


Fig. 19. Hotspot Mitigation Protocol Illustration

4.2.5.3 Evaluation of HMP

We have implemented HMP using the ns-2 simulator and its wireless extension. Full details can be found in [5]. The HMP implementation includes monitoring modules, measurement mechanisms, NST module, and the HMP algorithms discussed earlier. The simulated network size is 1200 meters x 1200 meters where 100 mobile nodes create 20 ~ 40 TCP and CBR/UDP flows. All data packets are a fixed size of 128 bytes and each simulation run lasts for 300 seconds unless specified otherwise. Each mobile node has a transmission range of 250 meters and shares a 2 Mbps radio channel with its neighboring mobile nodes. The simulations also include a two-ray ground reflection model, finite energy module, and IEEE 802.11 MAC protocol. During the evaluation, we use the terms 'HMP system' and 'baseline system' to refer to wireless ad-hoc networks with and without the HMP mechanisms, respectively. Both systems include the standard DSR and ADOV protocols.

We first observe how HMP systems perform compared to the baseline system in terms of packet delivery ratio (PDR). Figure 20 illustrates the comparison of packet delivery ratios for two differently configured HMP systems and the baseline system for increasing load. The two HMP systems are simply called HMP-P and HMP-R where HMP-R is more aggressive than HMP-P in its route suppression mechanism. HMP-P stands for HMP-POC where HMP mechanisms are executed only at points of congestion (POC). On the other hand, HMP-R represents HMP-Regional signifying its regional execution of hotspot mitigation algorithms. In other words, when a hotspot is detected HMP-P executes hotspot mitigation algorithms at the point of hotspots whereas HMP-R executes its mechanisms on a hotspot region. A node belongs to a hotspot region if it is a hotspot or it is an immediate neighbor of a hotspot. We note that both `enough_nh_neighbor` and `path_indicator` are always considered in all hotspot mitigation decisions.

As observed in Figure 20, HMP-P and HMP-R have little impact on lightly loaded network, e.g., below 100 Kbps. This is because the baseline system already achieves more than 90 % PDR and HMP has little room to make any improvements. However, as offered load increases, and

congestion builds up, HMP begins to provide improvements, as illustrated in the figure. Both HMP-P and HMP-R provide substantial improvements in the PDR. Specifically, HMP-P and HMP-R provide up to 43% and 46% increase in packet delivery ratio when compared with the baseline system. From Figure 20, we also observe the behavior of HMP-R is more radical than that of HMP-P. When the offered load is moderately high, HMP-R often outperforms others but becomes less effective when offered load is light, e.g., below 250 Kbps. Moreover, the performance of HMP-R varies with different loads, as illustrated in Figure 20. We conclude that HMP-R is too aggressive for lightly loaded networks rendering it only useful in heavily loaded networks.

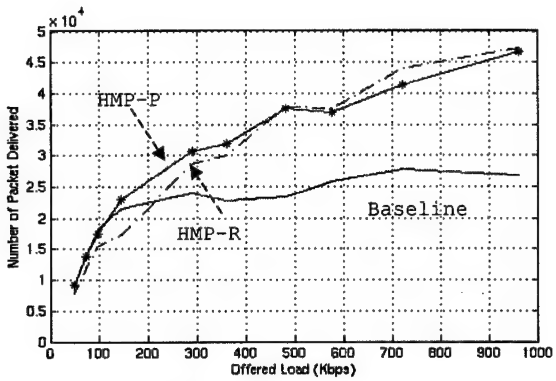


Fig. 20. Number of Data Packets Delivered

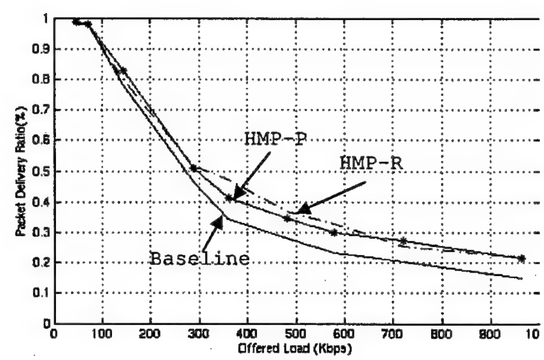


Fig. 21 Comparison of PDR against network load

Further details on HMP-P, HMP-R and baseline system are readily perceived through the number of delivered packets comparison among these systems, as illustrated in Figure 21. One interesting observation is that number of packets delivered in the baseline system levels-off around 2.3×10^4 but in HMP-P and HMP-R systems the number of delivered packets continuously increases with increasing offered load. There are two major reasons for this improvement. First, HMP creates route on non-congested nodes when possible allowing networks to utilize more distributed routes even if it means not the shortest path. Creating routes at non-hotspot nodes helps traversing flows to encounter fewer problems, and as a consequence, more packets are delivered. Second, HMP generates less routing overhead through suppression executed at hotspots. Many hotspot nodes rebroadcast route request packets and these route request packets often flood large areas of the network or even the entire network. However, many of these rebroadcast route request packets are lost before reaching their destination nodes. We observed that a considerable amount of route request packets are just wasted in the network without successful route creation in heavily loaded networks.

In HMP systems, routing packets (i.e., route request) are pre-filtered at hotspot nodes/regions. This not only prevents new routes being created through hotspots but also reduces 'to-be-lost' route requested packets that rely on broadcast/flooding. This opens up room for more data packets and as consequence more packets are delivered in HMP systems compared with that of baseline system. Moreover, as congestion conditions become more aggravated more nodes encounter packet loss and often interpret this packet loss as route errors, triggering numerous route recovery routines. As a consequence, additional routing overhead is added to an already congested network. However, in HMP systems congested nodes avoid participating in new route creations to mitigate their congested conditions, and consequently less routing packets are observed in the network.

We observe that HMP-R outperforms HMP-P when the offered load is excessive. However, HMP-R is too aggressive for lightly loaded network because we observe that the PDR of HMP-R

is less than that of HMP-P and no better than that of the baseline system when the offered load is less than 150 Kbps. However, both HMP systems outperform the baseline system.

5. List of all Publications and Technical Reports

Manuscripts submitted, but not published

1. Sun, L.-H. and Schwartz M., "Two-Tier Cellular System Channel Assignment Policies", *IEEE Journal on Selected Areas in Communications*, 2001, under submission

2. L. Sun and Mischa Schwartz, "Service Differentiation with a p-persistent Wireless MAC Protocol", 2001, under submission.

3. L. Sun and Mischa Schwartz, "Comparison of Call Admission and Packet Scheduling Policies for a Voice-Data Integrated Cellular System", Nov. 30, 2001, under submission.

4. Liao, R. R.-F. and Andrew T. Campbell, "Utility-based Network Adaptation for Multimedia Content Delivery", submitted to *IEEE/ACM Transactions on Networking*, August 2001

5. Lee, S-B. and Andrew T. Campbell, "Load Balancing in Mobile Ad Hoc Networks", submitted to *ACM MobiHoc 2003, November, 2002*

Papers published in peer-reviewed journals

6. Lee, S.B., Ahn G-S., Zhang, X., and A.T. Campbell, "INSIGNIA: An IP-Based Quality of Service Framework for Mobile Ad Hoc Networks", *Journal of Parallel and Distributed Computing (Academic Press)*, Special issue on Wireless and Mobile Computing and Communications, Vol. 60 No. 4 pg. 374-406, April 2000.

7. Epstein, Bracha and Schwartz, M., "Predictive QoS-Based Admission Control for Multi-class Traffic in Cellular Wireless Networks", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pg. 523- 534. March 2000.

8. Lee, S.B., Gahng-Seop, A., Zhang, X., and A.T. Campbell, "Improving TCP and UDP Performance in Mobile Ad Hoc Networks", *IEEE Communications Magazine*, Feature Topic on Challenges in Mobile Ad hoc Networks Vol. 39, No. 6, pp. 156-165, June 2001.

9. R.-F. Liao and A. T. Campbell, "A Utility-Based Approach to Quantitative Adaptation in Wireless Packet Networks", ACM Journal on Wireless Networks (WINET), Vol. 7, No. 5, pp. 541-557, September 2001.

10. Veres, Campbell, A.T, Barry, M. and L-H. Sun, "Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control", *IEEE Journal of Selected Areas in Communications (JSAC)*, Special Issue on Mobility and Resource Management in Next-Generation Wireless Systems, Vol. 19, No. 10, pp. 2094-2104, October 2001.

11. G.-S. Ahn, A. T. Campbell, Andras Veres and Li-Hsiang Sun, "Supporting Service Differentiation for Real-Time and Best Effort Traffic in Stateless Wireless Ad Hoc Networks (SWAN)", IEEE Transactions on Mobile Computing, (Special Issue of Best Wireless Papers from IEEE INFOCOM 2002), September 2002.

Papers published in non-peer-reviewed journals or in conference proceedings

12. Lee, S.B., and Andrew T. Campbell, "QOS Issues in Mobile Ad Hoc Networks", *Seminar on Multimedia Communications - Networks and Systems*, Invited Talk, Dagstuhl, Germany, May 4-6, 1999

13. Lee, S.B., Ahn, G-S., Zhang, X., and A.T. Campbell, "Evaluation of the INSIGNIA Signaling System" *8th IFIP International Conference on High Performance Networking* (Network 2000), Paris, France, May, 2000.

14. Campbell, Andrew T., "PARO: Power-Aware Routing in Ad Hoc Networks", *NSF/ONR Workshop on Cross-Layer Design in Adaptive Ad Hoc Networks: From Signal Processing to Global Networking*, Ithaca, New York, May 31-June 1, 2001

15. Barry, M., Campbell, A.T and A. Veres, "Distributed Control Algorithms for Service Differentiation in Wireless Packet Networks", Proc. IEEE INFOCOM'2001, Anchorage, Alaska, April 2001

16. Gahng-Seop Ahn, Andrew T. Campbell, Andras Veres and Li-Hsiang Sun, " SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks", Proc. IEEE INFOCOM'2002, New York, New York, June 23-27, 2002 (Selected as one of the six best wireless papers for a special issue of IEEE Transactions on Mobile Computing).

Papers presented at meetings, but not published in conference proceedings

17. Seoung-Bum Lee, Gahng-Seop Ahn, Xiaowei Zhang and Andrew T. Campbell "INSIGNIA" draft-ietf-manet-insignia-01.txt, *Internet-Draft*, Work in Progress, Presented at IETF MANET Working Group Meeting, Washington DC, December 1999.

18. Gahng-Seop Ahn, Andrew T. Campbell, Andras Veres and Li-Hsiang Sun "SWAN" draft-ietf-manet-swan-01.txt, *Internet-Draft*, Work in Progress, submitted for presentation, IETF MANET Working Group Meeting.

Publicly available software (source code) from the project

19. INSIGNIA, ns-2 extensions and tested code for QOS in Mobile Ad Hoc Networks Released: July 2000 Web: http://comet.columbia.edu/insignia/

20. SWAN, ns-2 and testbed code for QOS in Stateless Wireless Ad Hoc Networks Released: November 2002 Web: http://comet.columbia.edu/swan

Protocol project web pages

21. INSIGNIA Project Web Page: http://comet.columbia.edu/insignia/

22. SWAN Project Web Page: http://comet.columbia.edu/swan

23. HMP Project Web Page: http://comet.columbia.edu/hmp

6. List of all Participating Scientific Personnel

Professors: Andrew T. Campbell and Mischa Schwartz

Graduate Research Assistants: Seoung-Bum Lee, Gahng-Seop Ahn, Xiaowei Zhang, Raymond Liao and Li-Hsiang Sun

PhD Thesis Awards:

Li-Hsiang Sun, *"Bandwidth Allocation Schemes in Cellular and Wireless Local Area Networks"*, December 2001

Raymond R.-F. Liao, *"Utility-Based Adaptation, Dynamic Provisioning and Incentive Engineering Techniques for Internet and its Wireless Extensions*", December 2002

Visitors that worked on the project: Andras Veres (Ericsson Research)

7. Report on Inventions

1. INSIGNIA Protocol [21]
2. SWAN Protocol [22]
3. Hotspot Mitigation Protocol [23]

8. Bibliography

S.-B. Lee, G.-S. Ahn, X. Zhang, and A.T. Campbell, "INSIGNIA: An IP-Based Quality of Service Framework for Mobile Ad Hoc Networks," J. Parallel and Distributed Computing, special issue on wireless and mobile computing and communications, vol. 60, no. 4, pp. 374-406, Apr. 2000.

A. Veres, A.T. Campbell, M. Barry, and L.-H. Sun, "Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control," IEEE J. Selected Areas in Comm., special issue on mobility and resource management in next-generation wireless systems, vol. 19, no. 10, pp. 2094-2104, Oct. 2001.

J.L. Sobrinho and A.S. Krishnakumar, "Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Networks," IEEE J. Selected Areas in Comm., vol. 17, no. 8, pp. 1353-1368, Aug. 1999.

P. Sinha, N. Venkitaraman, R. Sivakumar, and V. Bharghavan, "WTCP: A ReliableTransport Protocol for Wireless Wide-Area Networks," Proc. ACM ACM/IEEE Int'l Conf. Mobile Computing and Networking (MobiCom), Aug. 1999.

K. Ramakrishnan, S. Floyd, and D. Black, "An Addition of Explicit Congestion Notification (ECN) to IP," Internet Draft, <draft-ietf-tsvwg- ecn-03.txt>, work-in-progress, Mar. 2001.

S. Kalyanaraman, D. Harrison, S. Arora, K. Wanglee, and G. Guarriello, "A One-Bit Feedback Enhanced Differentiated Services Architecture," Internet Draft, <draft-shivkuma-ecn-diffserv-01. txt>, work-in-progress, Mar. 1998.

G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," IEEE J. Selected Areas in Comm., vol. 18, 2002.

D. Chiu and R. Jain, "Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks," Computer Networks, 1989.

S. Nanda, D.J. Goodman, and U. Timer, "Performance of PRMA: APacket Voice Protocol for Cellular Systems," IEEE Trans. Vehicular Technology, vol. 40, Aug. 1991.

D. Bansal and H. Balakrishnan, "TCP-Friendly Congestion Control for Real-Time Streaming Applications," Technical Report, MIT-LCS-TR-806, MIT Laboratory for Computer Science, May 2000.

C.E. Perkins and E.M. Royer, "Ad-Hoc On Demand Distance Vector Routing," Proc. IEEE Workshop Mobile Computing Systems and Applications, Feb. 1999.

J. Broch, D. Maltz, D. Johnson, Y.-C. Hu, and J. Jetcheva, "A Performance Comparison of Multihop Wireless Ad Hoc Network Routing Protocols," Proc. ACM/IEEE Int'l Conf. Mobile Computing and Networking (MobiCom), Oct. 1998.

Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standard 802.11, June 1999.

F. Kelly, P. Key, and S. Zachary, "Distributed Admission Control," IEEE J. Selected Areas in Comm., vol. 18, no. 12, Dec. 2000. Global mobile information systems program, available at http:// www.darpa.mil/ito/research/glomo/index.html, 2001.

C.R. Lin and M. Gerla, "A Distributed Architecture for Multi-media in a Multihop Dynamic Packet Radio Network," Proc. IEEE Globecom '95, pp. 1468-1472, Nov. 1995.

R. Ramanathan and M. Streenstrup, "Hierarchically-Organized, Multi-Hop Mobile Wireless Networksfor Quality-of-Service Support," available at ftp://ftp.bbn.com/pub/ramanath/ mmwn-paper.ps, 2002.

P. Sinha, R. Sivakumar, and V. Bharghavan, "CEDAR: A Core-Extraction Distributed Ad HocroutingAlgorithm," Proc. IEEE INFOCOM '99, Mar. 1999.

S. Chen and K. Nahrstedt, "Distributed Quality-of-Service Rout-ing in Ad Hoc Networks," IEEE J. Special Areas in Comm., vol. 17, no. 8, Aug. 1999.

Medium Access Control (MAC) Enhancements for Quality of Service (QoS), Draft Supplement to IEEE Standard 802.11, Nov. 2001.

S. Floyd, "TCP and Explicit Congestion Notification," ACM Computer Comm. Rev., vol. 24, no. 5, Oct. 1994.

A. Demers, S. Keshav, and S. Shenker, "Analysis and Simulation of a Fair Queueing Algorithm," Proc. SIGCOMM Symp. Comm. Architectures and Protocols, Sept. 1989.

R.H. Katz, "Adaptation and Mobility in Wireless Information Systems," IEEE Personal Comm. Magazine, vol. 1, no. 1, pp. 6-17, 1994.

SWAN Project: http://comet.columbia.edu/swan, 2002.

G.-S. Ahn, A.T. Campbell, A. Veres, and L.-H. Sun, "SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks," Internet Draft, <draft-ahn-swan-0.1.txt>, work-in-progress, Sept. 2002.

G.-S. Ahn, A.T. Campbell, A. Veres, and L.-H. Sun, "SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks," Proc. IEEE INFOCOM '02, 2002.

L. Breslau, E. Knightly, S. Shenker, I. Stoica, and H. Zhang, "Endpoint Admission Control : Architectural Issues and Perfor-mance," Proc. ACM SIGCOMM '00, Aug. 2000.

C. Centinkaya and E. Knightly, "Egress Admission Control," Proc. IEEE INFOCOM '00, 2000.

G. Bianchi, A. Capone, and C. Petrioli, "Throughput Analysis of End-to-End Measurement-Based Admission Control in IP," Proc. IEEE INFOCOM '00, 2000.

R.-F. Liao and Andrew. T. Campbell, "A Utility-Based Approach to Quantitative Adaptation in Wireless Packet Networks", *ACM Journal on Wireless Networks (WINET)*, Vol. 7, No. 5, pp. 541-557, September 2001

9 (A). Technology Transfer

We released the INSIGNIA signaling system as extensions to the *ns* simulator to the public domain in August 2000. The code interoperates with the main routing protocols discussed in the MANET Working Group (i.e., AODV, DSR and TORA) and can be downloaded from the INSIGNIA Project home page (comet.columbia.edu/insignia/). A number of research teams have used the system and extended code. There have been over 400 downloads of the code since its release in July 2000

We have been active within the IETF MANET working group, which was established with support from industry, ARO and other military agencies. We believe that QOS will become an important work item of the working group over the next few years and any proposed solution must be able to interwork with the routing protocols that are currently being standardized. It is our intent to feed our results, technology and insights into the IETF over the course of the project.

We built an experimental INSIGNIA testbed based on IEEE 802.11. The testbed implements the INSIGNIA signaling systems with the DSR routing protocol. The testbed was successful demonstrated at ACM MobiCom 2000, Boston, August 6-11, 2000. We plan to release the source code for the experimental testbed Summer 2001.

We have initiated work on a "power-aware" routing protocol called PARO capable of routing packets in ad hoc network based on power related optimization and device behavior criteria. This new activity was a direct result of insights gained from this project. The work on PARO is joint work with IBM Research.

Prof. Campbell presented invited seminars on the QOS research issues to be addressed in mobile ad hoc networks, with emphasis on the ARO-sponsored work to the following organizations Lucent Technologies, ATT Research, Third ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (keynote address), Internet Engineering Task Force, WINLAB, Rutgers University and Polytechnic University.

We released the INSIGNIA signaling system as extensions to the *ns* simulator to the public domain in August 2000. The code interoperates with the main routing protocols discussed in the MANET Working Group (i.e., AODV, DSR and TORA) and can be downloaded from the INSIGNIA Project home page (comet.columbia.edu/insignia/). A number of research teams have used the system and extended code.

We have been active within the IETF MANET working group, which was established with support from industry, ARO and other military agencies. We believe that QOS will become an important work item of the working group over the next few years and any proposed solution must be able to interwork with the routing protocols that are currently being standardized. It is our intent to continue to feed our results, technology and insights into the IETF over the remainder of the project.

We built an experimental INSIGNIA testbed based on IEEE 802.11. The testbed implements the INSIGNIA signaling systems with the DSR routing protocol. The testbed was successful demonstrated at ACM MobiCom 2000, Boston, August 6-11, 2000. We released the source code

for the experimental testbed August 2001 for DSR and AODV stacks. The testbed code has been used by a number of researchers

We will release the ns-2 SWAN code and the experimental wireless testbed code once fully tested.

Prof. Campbell presented invited seminars on the QOS research issues to be addressed in mobile ad hoc networks, with emphasis on the ARO-sponsored work to the following organizations Internet Engineering Task Force, Polytechnic University, IBM, Ericsson, Microsoft Research and the NSF/ONR Workshop on Cross-Layer Design in Adaptive Ad Hoc Networks.

Prof. Campbell is organizing a panel on "Post 9-11 Networking Challenges" at IFIP Networking 2002, Pisa, Italy, May 2002. The panel will discuss the uses of the technology developed in this ARO sponsored project.

---

9 (B). Honors and Awards

Prof. Andrew T. Campbell received the NSF Faculty Career Development (CAREER) Award (1999-2003) and the IBM University Partnership Faculty Award (1999).

Prof. Mischa Schwartz received the IEEE Third Millennium Medal and the Eminent Member Award of Eta Kappa Nu. The latter award has been given to only 93 individuals since it was first established in 1950.

Raymond Liao received an ACM SIGCOMM/NSF travel grant to attend ACM SIGCOMM 2000, Stockholm

Prof. Andrew T. Campbell was invited to join the editorial boards for Computer Networks, IEEE/ACM Transaction on Networking, ACM Wireless Networks and ACM SIGCOMM Computer Communication Review

Prof. Andrew T. Campbell was invited to be a technical program co-chair ACM MobiCom 2002.

Prof. Andrew T. Campbell joined the steering committee for the ACM Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc)

Prof. Andrew T. Campbell joined the executive committee for ACM SIGMOBILE and became its Information Director.

Prof. Andrew T. Campbell gave the keynote address at the Third ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Boston, August 2000. The talk presented results from the ARO-sponsored project.

Prof. Andrew T. Campbell was invited to join the editorial boards for Computer Networks, IEEE/ACM Transaction on Networking, IEEE Transaction on Mobile Computing, IEEE Wireless Communications Magazine, ACM Wireless Networks and ACM SIGCOMM Computer Communication Review

Prof. Andrew T. Campbell was elected to the Executive Committee of the ACM Special Interest Group on Mobile Computing and Communications (ACM SIGMOBILE) where he currently serves as its secretary.

Prof. Andrew T. Campbell gave the keynote address at the Third ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Boston, August 2000.

Prof. Andrew T. Campbell gave the keynote address at the 6th International Conference on Protocols for Multimedia Systems (PROM 2001), October 17-19, 2001, Enschede, Netherlands. He also gave the keynote address at the Conference on the Path to 4G, Helsinki, 13-14 September 2001. The talks presented results from this ARO-sponsored project.

Prof. Andrew T. Campbell was invited to report on results from this ARO sponsored project at the NSF/ONR Workshop on Cross-Layer Design in Adaptive Ad Hoc Networks: From Signal Processing to Global Networking, Cornell University Ithaca, New York, May 31-June 1, 2001

Prof. Campbell served as program co-chair for the 4th IEEE International Conference on Open Architecture and Network Programming (OPENARCH 2001) and currently serves as technical program co-chair for the 8th ACM International Conference on Mobile Computing and Networking (ACM MobiCom 2002), and technical chair of the special track on networking technologies, services and protocols for IFIP Networking 2002.

Professor Campbell was the Technical Program Co-Chair for ACM MobiCom 2002.

Prof. Campbell is organizing a panel on "Post 9-11 Networking Challenges" at IFIP Networking 2002, Pisa, Italy, May 2002 with panelist: Randy Katz (University of California, Berkeley), Jonathan Liebenau (London School of Economics), Nicholas F. Maxemchuck (Columbia University), Karl Rauscher (Lucent, Founder Wireless Emergency Response Team). The panel will discuss the use of MANET and QOS capable MANETs in times of national disaster.